universität wien

The Proposal for an Artificial Intelligence Act from a consumer policy perspective

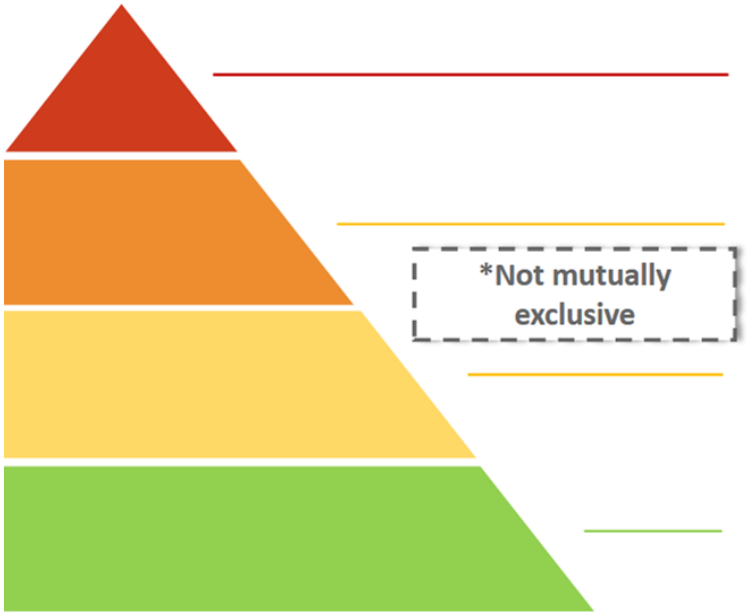
Christiane Wendehorst

Federal Ministry
Republic of Austria
Social Affairs, Health, Care and Consumer Protection

1. *Per se*-prohibited AI practices
2. High-risk AI systems
3. Individual rights
4. Enforcement

Per se-prohibited AI practices

The risk-based approach



Unacceptable risk
e.g. social scoring

Prohibited

High risk
e.g. recruitment, medical devices

Permitted subject to compliance with AI requirements and ex-ante conformity assessment

*Not mutually exclusive

'Transparency' risk
'Impersonation' (bots)

Permitted but subject to information/transparency obligations

Minimal or no risk

Permitted with no restrictions

European Commission

Manipulation and exploitation

universität wien

TITLE II

PROHIBITED ARTIFICIAL INTELLIGENCE PRACTICES

Article 5

1. The following artificial intelligence practices shall be prohibited:

(a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;

(b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;

(c) the placing on the market, putting into service or use of AI systems by public

What about mere economic harm?

What about other vulnerabilities?

Recommendations to rephrase

(b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of

(i) a specific group of persons due to their age, physical or men[illegible] disability **or social or economic situation; or**

(ii) **an individual whose vulnerabilities are characteristic of that individual's known or predicted personality or social or economic situation**

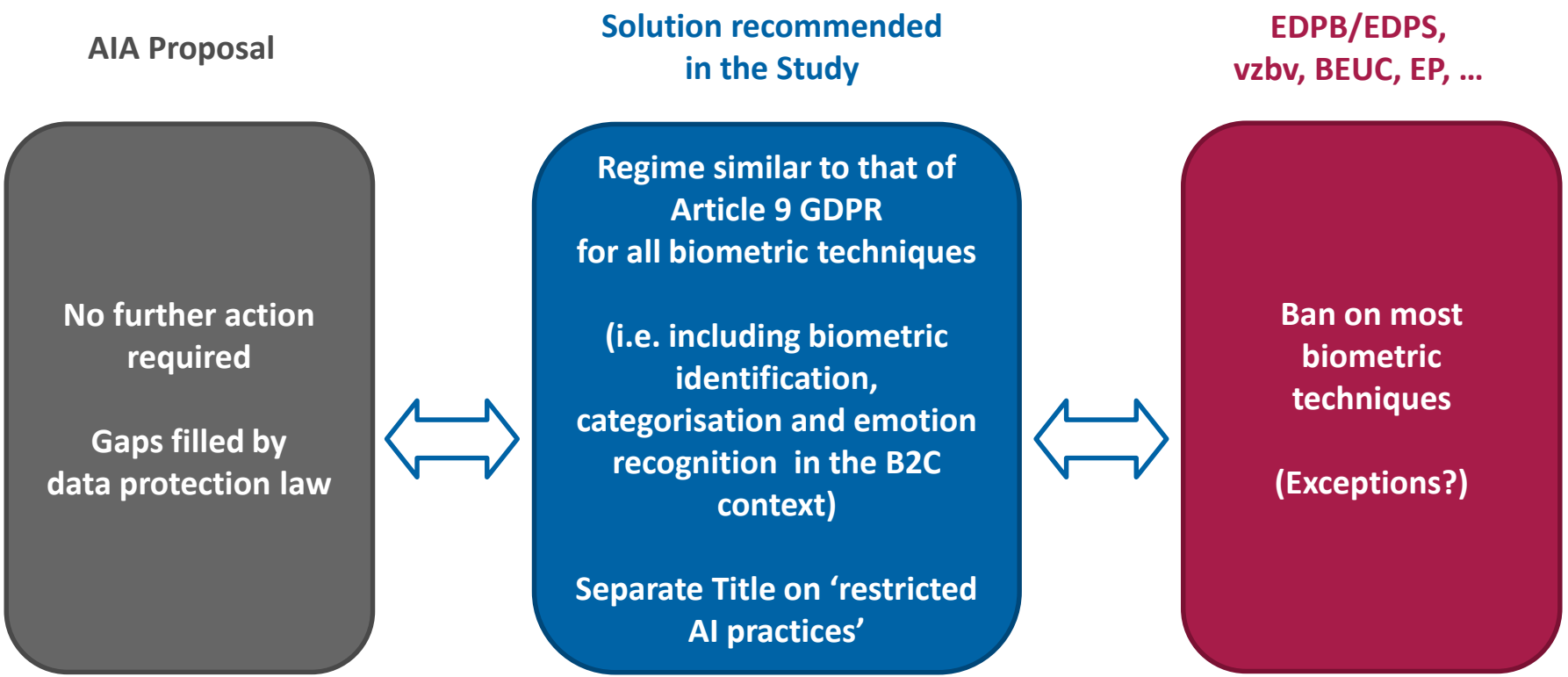
~~[illegible]der to materially distort the behaviour of a person pertaining to that group~~ in a [illegible] causes or is likely to cause that person or another person ~~physical or psychological~~ **material and unjustified** harm;

Include individual vulnerabilities and social / economic vulnerabilties

Include any infliction of significant harm

Similar recommendations to rephrase also for other items on the list

Biometric Techniques



Is the list in Article 5 complete?

Add prohibition of total surveillance

(ba) the putting into service or use of an AI system for the com[illegible]ehensive surveillance of natural persons in their private or work life to an extent or in a manner that causes or is likely to cause those persons or another person material and unjustified harm;

(bb) the placing on the market, putting into service or use of an AI system for the specific technical processing of brain data in order to read or manipulate a person's thoughts against that person's will or in a manner that causes or is likely to cause that person or another person material and unjustified harm.

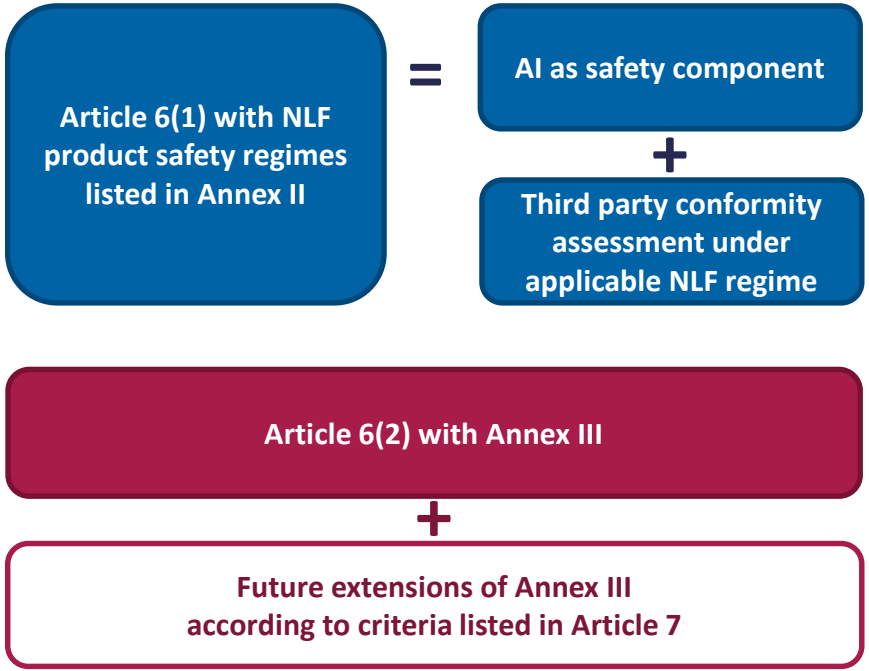
Add prohibition of certain processing of brain data

Recommendation to add also a flexibility clause and reference to prohibitions following from other law

universität wien

High-risk AI systems

What qualifies as high-risk AI system?



Usually leads to reasonable results with regard to traditional 'safety risks' of consumer goods (e.g. machinery), but not with regard to 'fundamental rights risks' (e.g. speaking doll, online game)

Covers only credit scoring in terms of 'consumer interests' in the narrower sense (not, e.g., risk assessment by insurance companies, personalised pricing, etc.), does not include any consumer goods

Pure economic risks etc

Clarify that the notion of 'fundamental rights risks' includes, inter alia, pure economic risks and risks for society at large

[…] Whereas: […]

(32) As regards stand-alone AI systems, meaning high-risk AI systems other than those that are safety components of products, or which are themselves products, it is appropriate to classify them as high-risk if, in the light of their intended purpose, they pose a high risk of harm to the health and safety or the fundamental rights of persons, taking into account both the severity of the possible harm and its probability of occurrence and they are used in a number of specifically pre-defined areas specified in the Regulation. The identification of those systems is based on the same methodology and criteria envisaged also for any future amendments of the list of high-risk AI systems. **The notion of fundamental rights risks is understood very broadly and not restricted to risks for rights explicitly mentioned under a separate Article in the Charter as long as there is a sufficient link between the risk and enjoyment of rights under the Charter. Notably, the notion of fundamental rights risk may, depending on the context, include mere economic risks where those risks are sufficiently severe, for instance because they affect access to essential goods or services (such as energy supply, credit or insurance) or because they operate on a large scale and may significantly affect the general standard of living of a natural person (such as large scale personalised pricing). The notion of fundamental rights risks also includes risks for society at large, such as for democratic institutions and a fair and open discourse.**

[…]

AI practices in the B2C context

5. Access to and enjoyment of essential private services and public services and benefits**, including access to products**:

(a) AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services;

(b) AI systems intended to be used

(i) to evaluate the creditworthiness of natural persons or establish their credit score~~,~~

(ii) **to evaluate the behaviour of natural persons such as with regard to complaints or the exercise of statutory or contractual rights in order to draw conclusions for their future access to private or public services,**

(iii) **for making individual risk assessments of natural persons in the context of access to essential private and public services, including insurance contracts, or**

(iv) **for personalised pricing within the meaning of Article 6 (1) (ea) of Directive 2011/83/EU,**

with the exception of AI systems put into service by small scale providers **of AI systems** for their own use;

Add practices such as individual risk assessment in the insurance context or personalised pricing

AI systems to be used by consumers

universität wien

5a. Use by vulnerable groups or in situations that imply vulnerability to fundamental rights risks

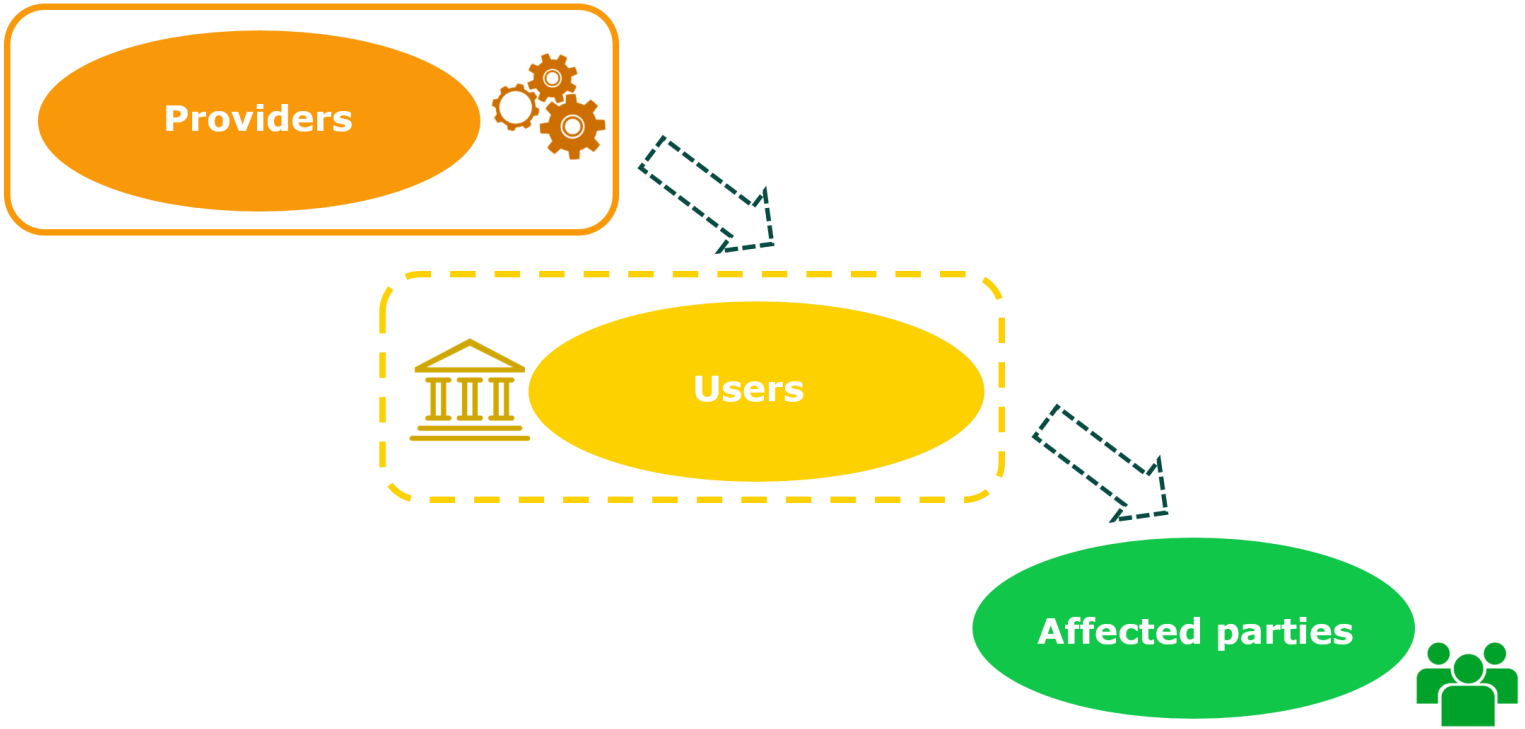
(a) AI systems intended to be used by children in a way that may seriously affect a child's personal development, such as by educating the child in a broad range of areas not limited to areas which parents or guardians can reasonably foresee at the time of the purchase;

(b) AI systems, such as virtual assistants, intended to be used by natur persons for taking decisions with regard to their private lives that ha legal effects or similarly significantly affect the natural persons;

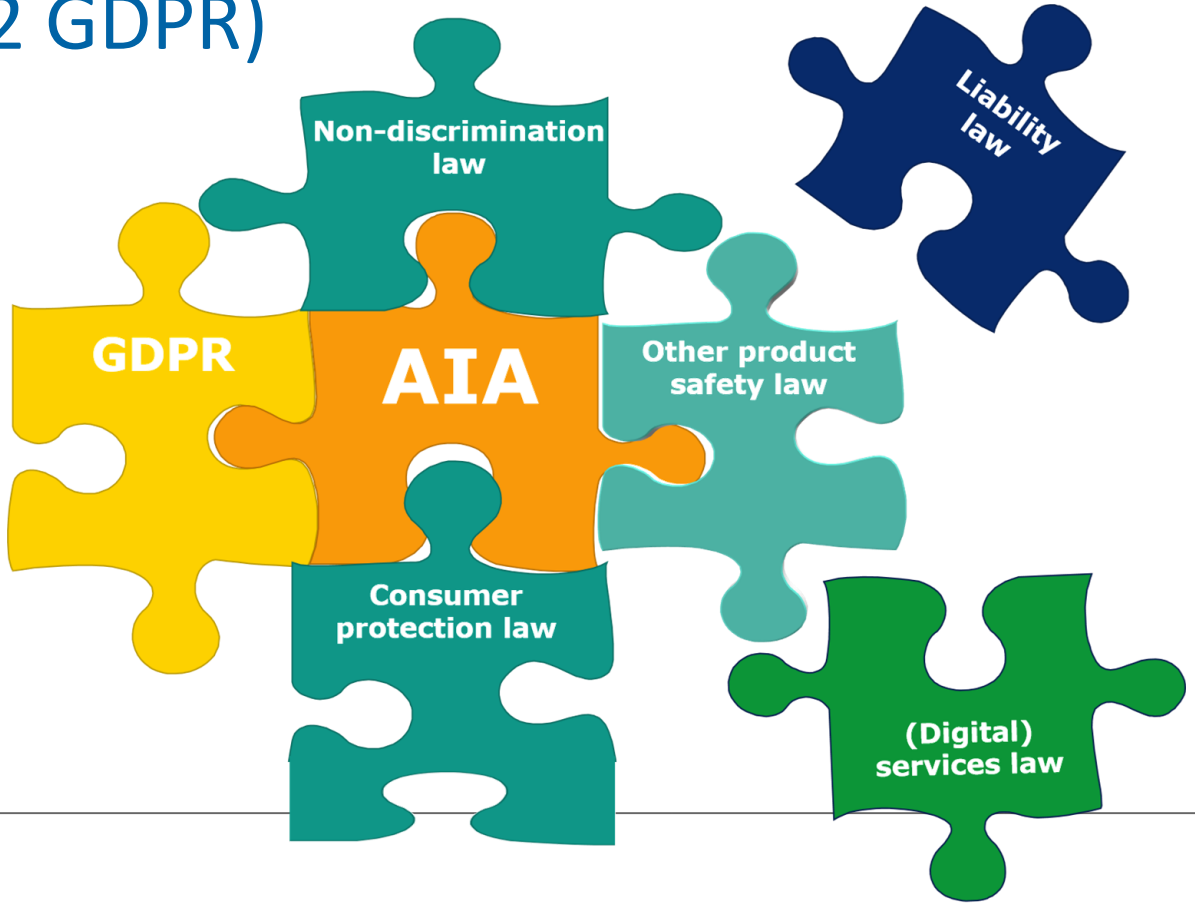
Add AI systems used by consumers themselves

3 Individual rights

The product safety law approach



Reliance on other laws (e.g. Art 22 GDPR)



Include fairness risks in Title IV

universität wien

TITLE IV

~~TRANSPARENCY~~ OBLIGATIONS FOR CERTAIN AI SYSTEMS POSING TRANSPARENCY OR FAIRNESS RISKS

Article 51a
Compliance with the obligations

1. This Title includes obligations for AI systems where one or both of the following conditions are fulfilled:

   (a) use of the AI system involves a risk of confusion between AI system and humans, or their operations or activities, where such confusion might harm the legitimate interests of persons exposed to the AI system;

   (b) use of the AI system leads to a decision with regard to a person that involves a material degree of evaluation or discretion and thus involves a fairness risk for the affected person.

2. The obligations of users of AI systems under this Title shall apply also to users who do not operate the AI system under their own authority but who solicit the

Transparency risks (already covered by proposed Title IV)

Fairness risks (suggested to be included in Title IV)

Scrutiny of individual decisions

Extension of the rule in Article 22 GDPR to AI-based decisions with meaningful human involvement

Creating more flexibility in order to allow for innovative solutions

Article 52a
Scrutiny of individual decision-making

1. No decision which produces legal effects concerning a person, or which similarly significantly affects that person, is taken by the user on the basis of the output from an AI system unless the appropriateness and fairness of this decision has been verified by means that are appropriate to the nature and the significance of the decision and the role of the AI system in the decision-making process.

2. Unless otherwise specified by Union or Member State law, verification within the meaning of paragraph 1 may, in particular, consist in meaningful scrutiny, before the decision is taken, by a natural person who is equipped with the appropriate

   (a) abilities, training and decision-making authority;

   (b) information with regard to the individual case; and

   (c) safeguards against automation bias.

3. The user may replace ex-ante verification within the meaning of paragraphs 1 and 2 by equivalent other measures where the affected person has given explicit consent or where ex-ante verification is impossible or would cause unreasonable effort and is not strictly necessary for safeguarding the affected person's rights and freedoms and legitimate interests. Unless otherwise specified by Union or Member State law, such equivalent other measures may, in particular, consist in the right to

   (a) obtain human intervention that satisfies the requirements under paragraph 2;

   (b) provide additional information and express his or her point of view; and

   (c) contest the decision with a meaningful chance of having it revised.

Explanation of individual decisions

Right to receive an explanation (main parameters, relative weight etc)

inspired by Articles 22 with 13-15 GDPR but is better adapted to the affected individual's needs and the needs of innovative business models

Article 52b
Explanation of individual decision-making

1. A decision which is taken by the user on the basis of the output from an AI system and which produces legal effects concerning a person, or which similarly significantly affects that person, shall be accompanied by a meaningful explanation of

   (a) the role of the AI system in the decision-making process;

   (b) the logic involved, the main parameters of decision-making, and their relative weight; and

   (c) the input data relating to the affected person and each of the main parameters on the basis of which the decision was made.

   For information on input data under point (c) to be meaningful it must include an easily understandable description of inferences drawn from other data if it is the inference that relates to a main parameter.

2. Paragraph 1 shall not apply to the use of AI systems

   (a) that have only minor influence within the decision-making process;

   (b) that are authorised by law to detect, prevent, investigate and prosecute criminal offences or other unlawful behaviour;

   (c) for which exceptions from, or restrictions to, the obligation under paragraph 1 follow from Union or Member State law, which lays down appropriate other safeguards for the affected person's rights and freedoms and legitimate interests; or

   (d) where the affected person has given explicit consent not to receive an explanation.

3. The explanation within the meaning of paragraph 1 shall be provided at the time when the decision is communicated to the affected person. However, the user may provide the explanation only at a later point upon the affected person's request, where providing the explanation immediately is not strictly necessary for safeguarding the affected person's rights and freedoms and legitimate interests, in which case the user shall inform the affected person of the right under this Article and how it can be exercised.

4 Enforcement

Inclusion of AIA in the Annex to the RAD

New Chapter on systemic risks

Obligations for very large providers to manage systemic risks

CHAPTER 2A

ADDITIONAL OBLIGATIONS FOR VERY LARGE PROVIDERS TO MANAGE SYSTEMIC RISKS

Article 62a
Very large providers

1. This Chapter shall apply to providers of high-risk AI systems listed in Annex III for which both of the following conditions are fulfilled:

   (a) the provider has a share of […] percent or above in the market for AI systems of the relevant type, considering the AI system's core functionalities, with regard to the whole Union, or a share of […] percent or above in the relevant market in at least three Member States; and

   (b) […] percent or above of decision-making of the relevant kind listed in Annex III significantly relies on the use of that type of AI system.

   When calculating the share within the meaning of point (a), AI systems that are not placed on the market or put into service under the provider's own name or trademark, but that use the provider's AI system as a basis or component in a way that significantly influences any systemic risks presented by those AI systems, shall be included.

2. The Commission shall adopt delegated acts in accordance with Articles 73 and 74, after consulting the Board, to lay down a specific methodology for calculating the market share referred to in paragraph 1. In those delegated acts, the Commission may also define different percentages than referred to in paragraph 1 for particular high-risk AI systems where there is reason to believe that systemic risks resulting from that type of AI system are significantly higher or lower than for other AI systems listed in Annex III.

New Chapter on systemic risks

- Systemic risk assessment
- Mitigation of systemic risks
- Independent audit
- Transparency reporting
- Data access and scrutiny by vetted researchers

Article 62b
Systemic risk assessment

1. As part of the quality management system referred to in Article 17 and post-market monitoring system referred to in Article 61, very large providers shall identify, analyse and assess, at least once a year, any significant systemic risks stemming from the functioning and use made of the AI systems provided by them in the Union.

2. This risk assessment shall be specific to the AI systems they provide and shall, in any case, include the following systemic risks:

   (a) any negative effects for the exercise of fundamental rights, for example respect for private and family life, data protection, the prohibition of discrimination, the rights of the child and access to an effective remedy and a fair trial, as enshrined in Articles 7, 8, 21, 24 and 47 of the Charter respectively;

   (b) any negative effects for democracy, the rule of law, the functioning of state institutions, the stability of societies and economies, protection of the environment and the combat against climate change, and other important public interests;

   (c) any risks resulting from uniformity of decision-making, including for the emergence of new disadvantaged groups, the reduction of diversity in affected groups (e.g. recruited individuals), and a steering function for human behaviour as affected individuals adapt their behaviour to the parameters relied on by the AI system;

   (d) any risks resulting from a reduction in human skills and competences, including for the ability to detect and correct errors and to act independently of the AI system where the system is unavailable;

   (e) risks of intentional manipulation of their AI system, including by means of targeted inauthentic behaviour of affected persons, malicious interferene by third parites, or hybrid warfare, with an actual or foreseeable negative effect on important public or private interests.